

What is claimed is:

Sub B1
1. A method for retrieving information stored in a database, wherein the database includes a set of records, wherein each of the records in the set includes a name field that stores a name, the method comprising the steps of:

receiving a database query, wherein the query includes a query name;

determining the records in the set that are likely to match the query, wherein the step of determining the records in the set that are likely to match the query comprises the steps of selecting one of the records in the set and determining whether at least a portion of the name stored in the selected record's name field has a pronunciation that is equivalent to a pronunciation of at least a portion of the query name; and

for each record that is determined to likely match the query:

comparing at least a portion of the name included in the record's name field to at least a portion of the query name; and

determining a similarity measurement between the query name and the name stored in the record's name field based on the comparison.

2. The method of claim 1, wherein the step of comparing at least a portion of the name stored in the record's name field to at least a portion of the query name comprises the step of performing n-gram comparisons.

3. The method of claim 1, wherein the query name consists of one or more character strings, wherein each character string consists essentially of letters of the Roman alphabet.

4. The method of claim 3, wherein, for each record in the set, the method further comprises the steps of:

using symbols from a phonetic alphabet to generate a character string that represents a pronunciation of at least a portion of the name stored in the record's name field; and

associating the generated character string with the record.

5. The method of claim 4, further comprising the step of using symbols from the phonetic alphabet to generate at least one character string that represents a pronunciation of at least a portion of the query name.

6. The method of claim 5, wherein the step of determining whether at least a portion of the name stored in the selected record's name field has a pronunciation that is equivalent to a pronunciation of at least a portion of the query name comprises the step of comparing the generated character string that is associated with the record to the generated character string that represents a pronunciation of at least a portion of the query name.

7. The method of claim 4, further comprising the steps of:
using symbols from the phonetic alphabet to generate a first character string that represents a first pronunciation of at least a portion of the query name; and
using symbols from the phonetic alphabet to generate a second character string that represents a second pronunciation of said portion of the query name.

8. The method of claim 7, wherein the step of determining whether at least a portion of the name stored in the selected record's name field has a pronunciation that is equivalent to a pronunciation of at least a portion of the query name comprises the step of comparing the generated character string associated with the record to the first character string and/or the second character string.

9. The method of claim 1, wherein the query name is a full name.

10. The method of claim 1, wherein the query name is a first name.

11. The method of claim 1, wherein the query name is a surname.

12. The method of claim 1, wherein the query name comprises a first name and/or a surname.

13. The method of claim 1, wherein each of said name fields stores a first name and/or a surname.

14. A method for retrieving information stored in a database, wherein the database includes a set of records, wherein each record in the set includes a name field that stores a name, the method comprising the steps of:

receiving a database query, wherein the query includes a query name;

analyzing the query name to determine whether it belongs to a culture that is included in a set of identified cultures;

if the query name appears to belong to a culture that is included in the set of identified cultures, then selecting a set of rules and/or a set of algorithms that is associated with the culture to which the query name appears to belong, otherwise selecting a default set of rules and/or algorithms;

using at least a portion of the query name and a rule and/or algorithm from the selected set of rules and/or algorithms to generate one or more keys; and

determining those records in the set of records that match at least one of the generated keys.

15. The method of claim 14, further comprises the steps of:

selecting a record that was determined to match at least one of the generated keys; and

comparing at least a portion of the name stored in the record's name field to at least a portion of the query name to determine a similarity measurement between the query name and the name stored in the record's name field.

16. The method of claim 15, wherein the step of comparing at least a portion of the name stored in the record's name field to at least a portion of the query name comprises the step of performing n -gram comparisons.

17. The method of claim 14, wherein the step of determining the records in the set that match at least one of the generated keys comprises the step of determining whether a key that is associated with a record in the set matches at least one of the generated keys.

18. A method for determining whether an input string identifies a name that appears to belong to a culture that is included in a set of identified cultures, comprising the steps of:

- selecting a culture from the set of identified cultures;
- determining a surname that occurs with high frequency in the selected culture;
- determining a given name that occurs with high frequency in the selected culture;
- storing the surname in surname list, wherein each surname in the surname list is associated with one or more culture identifiers, and each of said one or more culture identifiers is associated with a confidence score; and
- storing the given name in a given name list, wherein each given name in the given name list is associated with one or more culture identifiers, and each of said one or more culture identifiers is associated with a confidence score.

19. The method of claim 18, further comprising the steps of:

- parsing the input string to identify a portion of the input string that appears to be a surname;
- parsing the input string to identify a portion of the input string that appears to be a given name;
- determining whether said portion of the string that appears to be a surname matches a surname included in said surname list; and
- determining whether said portion of the string that appears to be a given name matches a given name included in said given name list.

20. The method of claim 19, further comprising the step of:

- if said portion of the string that appears to be a surname matches a surname included in said surname list, then determining the one or more culture identifiers associated with said surname included in said surname list and determining the confidence score associated with each of said one or more culture identifiers; and

- if said portion of the string that appears to be a given name matches a given name included in said given name list, then determining the one or more culture identifiers

associated with said given name included in said given name list and determining the confidence score associated with each of said one or more culture identifiers.

21. The method of claim 20, further comprising the step of determining a confidence level that the name appears to belong to a particular culture, wherein the confidence level is a function of: (1) the confidence score associated with the culture identifier that (a) identifies the particular culture and that (b) is associated with said surname in said surname list that matched said portion of the input string that appears to be a surname, and (2) the confidence score associated with the culture identifier that (a) identifies the particular culture and that (b) is associated with said given name in said given name list that matched said portion of the input string that appears to be a given name.

22. The method of claim 21, further comprising the step of:
computing a likelihood that said portion of the string that appears to be a surname has a particular cultural origin based on a statistical model derived from digraph distribution statistics for names within various cultures.

23. The method of claim 18, further comprising the steps of:
determining a title, name affix, and/or name qualifier that occurs with high frequency in the selected culture;
storing the determined title, name affix, and/or name qualifier in a TAQ lookup table;
associating a culture identifier with the determined title, name affix, and/or name qualifier, wherein the culture identifier identifies the selected culture; and
associating a confidence score with the culture identifier associated with the determined title, name affix, and/or name qualifier.

24. The method of claim 23, further comprising the steps of:
segmenting the input string based on spaces in the input string; and
for each segment present in the input string, determining whether that segment matches a title, name affix, or name qualifier included in said TAQ lookup table.

25. The method of claim 18, further comprising the step of:
generating a morpheme data store comprising a plurality of morphemes, wherein each morpheme in the morpheme data store is associated with one or more culture identifiers, and each of the one or more culture identifiers is associated with a confidence score.

26. The method of claim 25, further comprising the step of determining whether one or more of the plurality of morphemes stored in the morpheme data store are present in the input string.

27. The method of claim 25, further comprising the step of:
generating an n-gram data store comprising a plurality of n-grams, wherein each n-gram in the n-gram data store is associated with one or more culture identifiers, and each of the one or more culture identifiers is associated with a confidence score.

28. The method of claim 27, further comprising the step of determining whether one or more of the plurality of n-grams stored in the n-gram data store are present in the input string.

29. A method for comparing a first proper name to a second proper name, comprising:
determining a culture to which the first proper name appears to belong;
selecting a rule and/or algorithm associated with said culture;
generating a plurality of keys based on the selected rule and/or algorithm and at least a portion of the first proper name; and
compare a key associated with the second name to one or more of the plurality of generated keys.

30. The method of claim 29, wherein if the keys match, compare at least a portion of the first proper name to at least a portion of the second proper name to determine a similarity measurement between the first proper name and the second proper name.

31. The method of claim 30, wherein the step of comparing at least a portion of the first proper name to at least a portion of the second proper name comprises the step of performing n-gram comparisons.

RES 58580v2

ACCESSION # 012502